

**Extended GIS Analysis for the  
Bay Area Travel Survey 2000 (BATS2000)**

**Prospectus**

**Planning Section  
Metropolitan Transportation Commission  
101 Eighth Street  
Oakland, CA 94607**

**December 2003**

## **I. Introduction**

The purpose of this prospectus is to outline current and future efforts at adding various geographic variables to the MTC regional household travel survey datasets. The Bay Area Travel Survey 2000 (BATS2000) was a \$1.5 million study to obtain detailed in-home and out-of-home travel and activity statistics from over 15,000 Bay Area households. BATS2000 is the fourth in a series of major travel surveys conducted in the Bay Area over the past forty years, starting with the original 1965 Bay Area Travel Survey; and continuing with other major surveys in 1981, 1990 and 1996. The BATS2000 dataset is the centerpiece for new sets of travel behavior models that will be produced over the next several years. The Bay Area Travel Survey also provides a comprehensive picture of regional and sub-regional travel characteristics, and is the only database that allows analysts to examine the full picture of both work and non-work travel patterns within the nine-county region. The final survey consultant report for the BATS2000, and early releases of the data files, are available on MTC's web page at: <http://www.mtc.ca.gov/datamart/survey.htm>.

In contrast with previous Bay Area household travel surveys, the BATS2000 retains detailed information on the precise home locations, and precise origins and destinations of all trips. This level of precision is a significant improvement to the older MTC surveys, which only retained census tracts or travel analysis zone information. This is a very important upgrade: new sets of MTC travel behavior models can be based on point-to-point travel times and distances, instead of zone-to-zone travel times and distances. As another example, household accessibility can be estimated from the actual location of the respondent household, instead of using zonal averages.

Current GIS activities with BATS2000 include the completion of geocoding, and re-geocoding of household locations and trip end locations to MTC's new street base map (GDT, Inc.) This includes the detailed x-coordinates and y-coordinates, in NAD83 – UTM – Zone 10N, meters; as well as standard Census Bureau and MTC geography including: block, block group, census tract, public use microdata area (PUMA), MTC 34 superdistrict, MTC 1454 travel analysis zone, and county. (Note that census geography has been adjusted by GDT to align with the GDT street layers.)

MTC has also conducted preliminary testing of these issues using ESRI's Network Analyst (for ArcView 3.x) and Spatial Analyst (for ArcView 3.x and ArcGIS 8.x) software. These tests are discussed in the following sections. The most significant challenge is: how to automate the GIS process, to apply in a "batch" mode, the many thousands if not hundreds of thousands of data points.

One of MTC's intents in issuing this prospectus is to garner interest in this project. This project has immediate applications to the San Francisco Bay Area, and potential applications to the scores of other household travel surveys conducted by state Departments of Transportation and Metropolitan Planning Organizations in the U.S. This project may appeal to software producers, such as ESRI, and possibly to advanced graduate students, PhD candidates and researchers in university-level geography, civil engineering and urban planning departments.

ESRI and the university community may be able to point out simple solutions to our problems; or create Avenue, VBA or other scripts that meet the needs of our project; or to produce future enhancements to existing ESRI products including Network Analyst or Spatial Analyst; or to

recommend other software products to use, such as ESRI's Net Engine, or the planned Network Analyst extension for ArcGIS 9.1; or to recommend professional consultants to contract for these improvements.

Some of these products, such as producing door-to-door distances for all trip records in the survey file, are straightforward but potentially very time-consuming in terms of CPU resources. This is our highest priority given that trip distance information will be very useful in our forthcoming (spring 2004) report on "Regional Travel Characteristics." We would like to solve our other issues before the end of calendar year 2004.

## **II. Trip File Analysis**

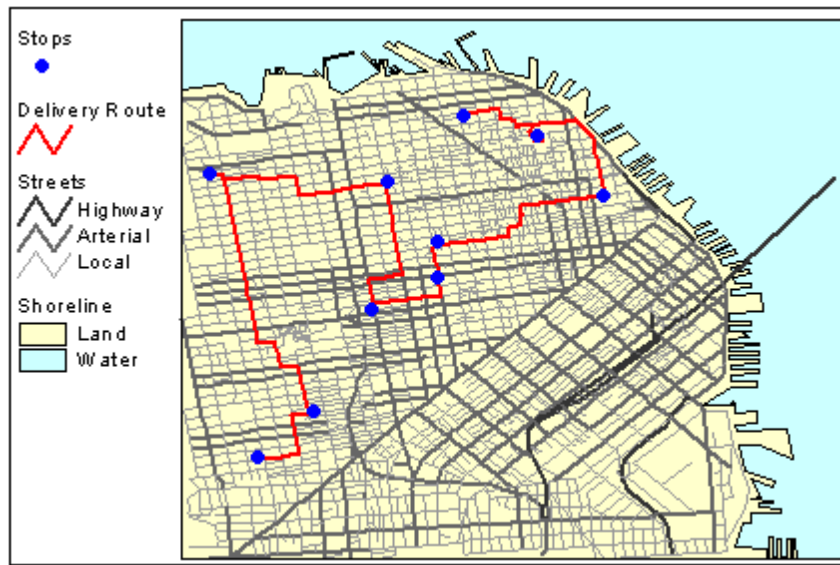
The BATS 2000 activity/trip file contains full information on in-home activities, out-of-home activities, and trips. Data is for 15,064 sample households for two consecutive days. There are over 259,000 individual trip records in the BATS database. The size of the file requires that all GIS analyses should be produced in a "batch" mode, not interactively.

### **1. Door-to-Door Distance**

The desire is to create polylines to represent every trip record in BATS2000. A minimum "cost" route would be associated with every trip record with valid origin and destination coordinates, and a distance (meters or miles) associated with every polyline.

MTC staff has successfully tested ESRI Network Analyst to create these door-to-door polylines. The process requires pre-processing and post-processing of the travel survey data using PC-SAS. SAS is used to re-assemble the source trip records, with the origin and destination on each record, to a consecutive "stops" file, with every origin and destination represented as a distinct record. The Network Analyst "Find Best Route" feature is then used to create a full set of polylines. This is still a kludged process, and yields extraneous polylines. For example, this process will create a polyline from the last destination of person "A" to the first origin of person "B." This "consecutive stops" database with the "Find Best Route" approach works well for purposes of designing delivery schedules, as shown in the following diagram. It would be useful for Network Analyst to have this "Find Best Route" feature work in instances where the origin and destination of a single trip is represented on a single data record.

### Example: “Find Best Route” option using Network Analyst



We will probably want several versions of the GDT network, including an “all highway” network and a “non-motorized” network (i.e., eliminating freeways from the path choice, with exceptions such as Golden Gate Bridge, Dumbarton Bridge, etc.) We will also want to work with several path-building options. For non-motorized (bicycle and walk) networks we will most likely need just a minimum distance path. For motorized paths we will need the distance for minimum impedance paths, e.g., assigning faster speeds to freeways, slower speeds for arterials, and extremely slow speeds for non-motorized trails and paths that may occur in the GDT networks.

The reasons for developing these door-to-door distances are:

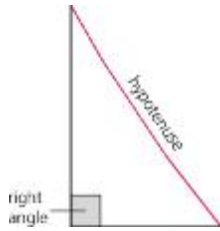
- 1) to check the reasonableness of reported travel times and computed travel speeds by mode;
- 2) to compute average trip lengths and trip length frequency distributions by purpose and mode;
- 3) to check the aggregate vehicle miles of travel from trip-based polylines against the two-day odometer VMT calculations, for all vehicles; and
- 4) for using the point-to-point distance and time values for estimating the non-motorized and perhaps motorized components of future mode choice models.

To provide some historical context, in the past decade we basically had to use zone-to-zone distances, from our “non-motorized” network, for use in model estimation and model application. With our detailed latitude/longitude databases and our GIS systems we will be much better served by calculating point-to-point distances for walk, bicycle and motorized trips.

## 2. Straight-Line Distance

This would also be known as “Pythagorean” or “Crow-Fly” distance since we would be computing the distance over the “hypotenuse” of a triangle, based on the detailed coordinates of the trip origins and trip destinations. This is a non-network-based approach for calculating distance.

Frankly we don't need the polylines for this sub-project, since we can basically calculate the hypotenuse distance using the UTM coordinates of the trip origins and trip destinations. These calculations can be done in either SAS or ArcView or ArcGIS. We may, however, be able to adapt the "desire line" extension to handle the point-to-point desire line production. (The "desire line" ArcView extension is currently programmed to create polylines from the centroid of all polygons to the centroid of all other polygons.)



### 3. X-Y Distance

We would also like to calculate the X-Y distance from origin to destination. This would be a mechanical calculation using SAS or ArcView. This would be the sum of the distances of the other two legs of a right triangle (opposite the hypotenuse). No polylines should be produced. The thought behind this calculation is that in a perfect, north-south gridded environment, the trip would be no shorter or no longer than these two triangle legs. In real life, however, the road distance will probably be longer than the straight-line (crow fly) distance as well as the x-y distance.

### 4. Door-to-Door Total Elevation Change

An extension to our creation of door-to-door polylines is a sub-project to calculate the total elevation gain and total elevation loss along all trip polylines. The reason for obtaining this information will be to understand the influence of topography on the choice for bicycling and walking. Trips with large (200'+) elevation gains and losses will typically have a higher disutility associated with walking and cycling.

MTC staff has successfully used ESRI Spatial Analyst to produce a polyline-based estimate of elevation changes. First, the trip polyline is converted, or "rasterized" into grid cells using Spatial Analyst. Spatial Analyst can then be used, in conjunction with topography layers, to ascertain the absolute elevation change between one grid cell to the next grid cell. The sum of these absolute elevation changes can then be associated to the original polyline, i.e., the original trip record. While MTC staff has explored this technique for a sample of random polylines, what we need is a "batch" process that will use a full set of trip records, converted into polylines, and use this Spatial Analyst or other process to obtain this net elevation change characteristic.

## **III. Household File Analysis**

There are 15,064 households in the final BATS2000 data.

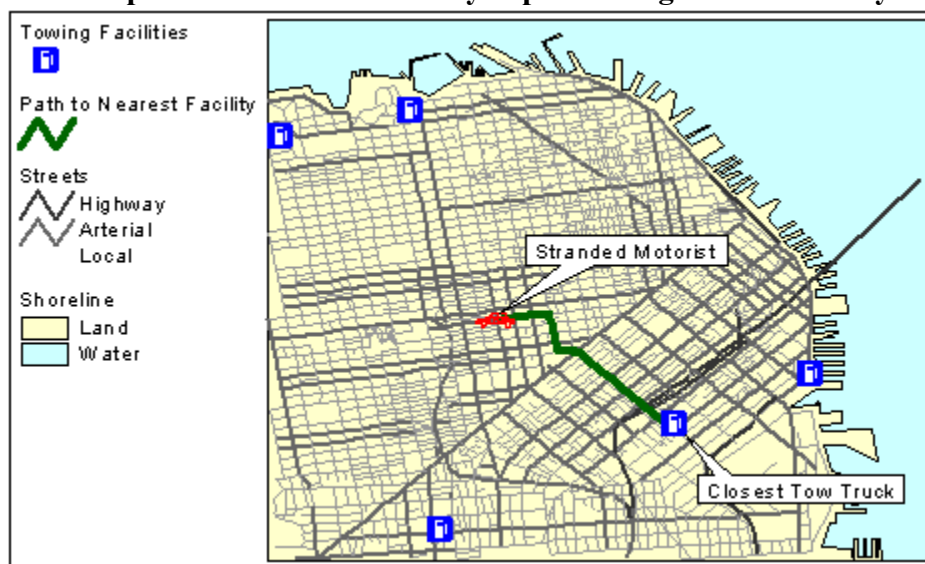
## 1. Walk distance to nearest transit stop

We need to know the walk accessibility of all BATS households to the nearest bus stop / rail station. This is needed in travel model estimation so that we can clearly understand each individual household's accessibility to transit. (If the nearest bus stop to a household is greater than 1.0 mile, then we know that this household is either going to drive to transit, or will not have transit available to them.)

This will be a challenge since we will need the complete regional database ("facility theme") of bus stops / rail stations. Current plans for a separate MTC project (the Regional Transit Information System) will result in the creation of a region-wide database of all bus and rail stations and stops, sometime in 2004.

We know about the ESRI Network Analyst "Find Closest Facility" option. This process uses an "event theme" (say, household locations) with a "facility theme" (say, bus stops / rail stations). This process creates a polyline from every event to the nearest facility. The challenge is to automate this into a "batch" process to handle all of our 15,000 events (households).

### **Example: "Find Closest Facility" option using Network Analyst**



## 2. Drive distance to nearest transit stop

Similar to walk accessibility-to-transit, we need to compute the minimum time and distance from all households to the nearest transit park-and-ride site. This will be a more limited set of destination points and will include all transit park-and-ride lots for BART, Caltrain, Ferries, Santa Clara LRT, and express buses-with-Park-and-Ride.

MTC staff will need to either work with the master regional transit database, or with MTC's TP+ transit networks, to extract a set of "transit park-and-ride sites" as a GIS point layer. The "park and ride" point layer is then snapped to the GDT base map, and the path-builder would build polylines from each of the 15,064 households to the nearest (distance and time-wise) park-and-ride node.

Given the incomplete status of the master regional transit database “station node layer” it may be prudent to begin on this sub-project before starting work on the walk accessibility to transit sub-project.

(We will probably not want to limit our “nearest transit stop” and “nearest transit park-and-ride” to the household records. We will probably want to apply these same procedures to every trip end in the trip record. We really do need to know how accessible all trips are to and from transit!)

### 3. Household-Level Neighborhood Density & Accessibility

We have typically used residential density and employment density in many of our travel behavior models. The typical use is the residential density of the zone-of-residence, or the employment density of the zone-of-residence. (We use the employment density as a measure of the shopping/working/leisure activities within close proximity, e.g., walking or bicycling distance, of the home of the traveler.)

We will want to create polygon buffers around all 15,064 households in BATS. These could be circle-shaped polygons with buffers (circle radius) of x, y and z distances. Perhaps ¼-mile, ½-mile and 1-mile radiuses, but we can experiment. We will then want to extract block level data from Census Summary File #1 (total population, total households, etc.); and census zone level data from the CTPP Part 2 file (e.g., workers by zone-of-work as a surrogate for total employment.) (The CTPP Part 2 data file is expected early 2004).

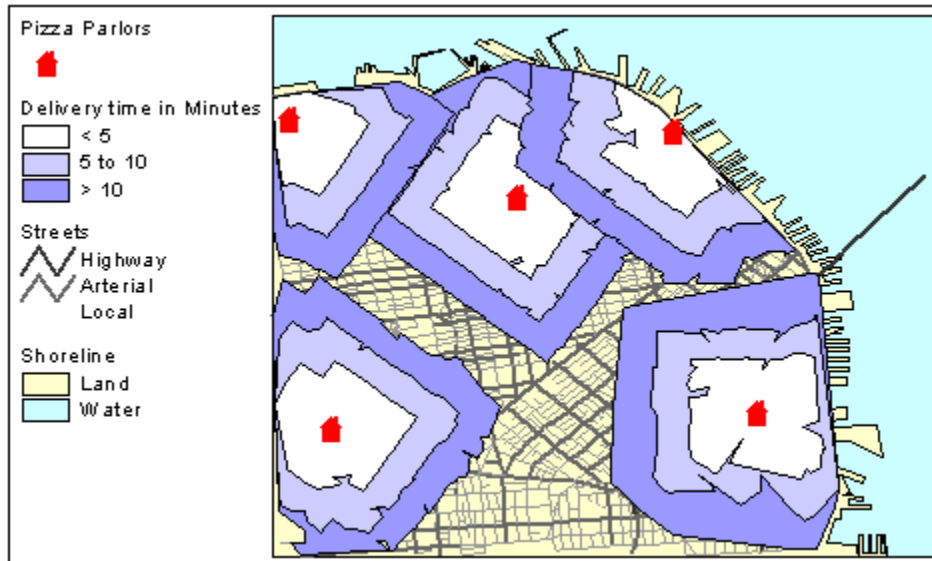
We can calculate the land area of the circle-polygon around each sample household, and each sample household will have it's own unique residential density, housing density, and employment density data values. (It will then be interesting to compare the disaggregate density approach to the traditional zone-of-residence-based density approach).

We searched the ESRI ArcScripts library for utilities to help automate this process and found the “Data Partitioner” script by Jieanwie Dou (8/8/02); and a “Calculate Demographics” script by Tim Johnson (2/21/01). These scripts or other utilities for ArcView 3.X or ArcGIS 8.X can be downloaded and tested.

Instead of circle-polygons surrounding households, we may develop network-based polygons using the “Finding Service Area” option of ESRI Network Analyst (see example below). This network-based approach makes more sense in that accessibility should be measured using real roads and networks as opposed to the fiction of traveling through solid buildings and across blocks of land (as assumed by the circle-polygon approach.)

Again, the challenge with either producing circle-shaped “buffer” polygons or network-based travel time (isochron) polygons based on network costs is the batch automation needed to produce this information for over 15,000 data points.

#### **Example: “Finding Service Area” option using Network Analyst**



#### **IV. Next Steps**

The next steps are for MTC staff to “shop around” this prospectus with ESRI and university colleagues. Our genuine hope is these are simple enough problems that will require minimal programming of new scripts for ArcView 3.X or ArcGIS 8.X, or are some other untested features of products that we are not aware of.

In terms of reporting results of this effort, we will want to produce summaries of all of this GIS-based “added value” to BATS2000, including:

1. Average trip lengths and trip length frequency distributions;
2. Average trip speeds and speed frequency distributions, by means of transportation;
3. Average gross elevation changes, and frequency distributions, by means of transportation and trip length;
4. Average distance, and frequency distribution of households, by closest transit facility;
5. Household characteristics by density characteristics;

In terms of MTC travel model research we will also want to test the sensitivity of travel model parameters to the use of “aggregate zonal-based” measures to the proposed use of “disaggregate GIS-based” measures.